(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 November 2002 (28.11.2002)

PCT

(10) International Publication Number
WO 02/096120 A1

(51) International Patent Classification[7]: H04N 7/50, 7/26

(21) International Application Number: PCT/SG01/00105

(22) International Filing Date: 25 May 2001 (25.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): CENTRE FOR SIGNAL PROCESSING, NANYANG TECHNOLOGICAL UNIVERSITY [SG/SG]; School of Electrical & Electronic Engineering, Block S2, Level B4, S2-B4b-05, 50 Nanyang Avenue, 639798 Singapore (SG).

(72) Inventors; and
(75) Inventors/Applicants (for US only): LI, Zhengguo [CN/SG]; c/o Centre for Signal Processing, Nanyang Technological University, School of Electrical & Electronic Engineering, Block S2, Level B4, S2-B4b-05, 50 Nanyang Avenue (SG). LIN, Xiao [SG/US]; c/o 22320 Foothill Blvd., Suite 460, Hayward, CA 94541 (US).

(74) Agents: NAMAZIE, Farah et al.; Haq & Namazie Partnership, Robinson Road Post Office, P.O. Box 765, 901515 Singapore (SG).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/096120 A1

(54) Title: BIT RATE CONTROL FOR VIDEO COMPRESSION

(57) Abstract: A bit rate control scheme for video encoding is described, in which a target bitrate for a picture frame (or video object or macroblock) is determined based ona fluid-flow model of the buffer dynamics, and in which the buffer targetoccupancy is set to about 50% of a buffer safety margin used to determinewhether a frame should be skipped or not, the margin being about 80% of the buffer size. A new Rate-Distortion Model for determining a suitable quantizationparameter to give the target bit rate is also described, as is a sliding windowmethod of determining prior data points to update the Rate-Distortion modelparameters, and a switching frame-skipping control which switches between a predictive skipping control and a post frame skipping control.

Bit Rate Control for Video Compression

Technical Field

The present invention relates to a bit rate control for the compression of video data. It has particular, but not exclusive, application to the provision of video over a packet switched network such as the Internet.

Background Art

Bit rate control plays an important role in the provision of multimedia over communications networks, and has been widely studied by many researchers for various standards and applications, such as storage media and real-time transmission with MPEG-1 and MPEG-2, videoconferencing with H.261 and H.263, and video object coding with MPEG-4.

For different coding standards and applications, different coding parameters are emphasised and different mechanisms are applied. For example, in MPEG-2, the most influential coding parameter with regard to picture quality is the quantization parameter (QP) used for texture coding. This parameter can be selected for an entire frame of the video sequence or can change from macroblock to macroblock. In most implementations, it is selected on the basis of buffer fullness, so that the buffer occupancy is maintained at a given level. The H.263 coding scheme allows for variable frameskip, and due to the low bit-rate conditions which may be imposed upon the encoder, it is up to the rate control algorithm to make appropriate decisions on both spatial and temporal coding parameters. If the buffer is in danger of overflow, complete frames may be disregarded at the encoder to allow bits used for the previous frame to be transmitted out of the buffer to thereby reduce the buffer level and delay. In conjunction with this frame-skipping mechanism, the bit rate control algorithm must determine a suitable quantization parameter (QP) to obtain the desired bit rate.

Similarly to H.263, MPEG-4 bit rate control also considers spatial and temporal coding parameters. However, the encoder must also consider the significant amount of bits which are used to code shape information such that arbitrarily shaped objects can be coded. Also, although each video object may be encoded at a different frame rate, it is preferable that all of the objects are encoded at the same frame rate in order to yield better video quality. Further, additional coding parameters are introduced by MPEG-4 to control the amount of bits used to specify the shape of an object. It is the responsibility of the rate control scheme to incorporate these new parameter decisions along with other parameter decisions to ensure that the video objects are effectively coded.

In real-time video communications, the encoded bits are placed into an encoder buffer before it is transmitted through a network to a decoder. If the actual bit rate of the encoder is greater than the available channel bandwidth, the additional bits accumulate in the encoder buffer and increase buffer delay, which is the time needed to send the buffer bits remaining from the previously encoded frames. When the number of bits in the buffer is too high, the encoder usually skips some frames to reduce the buffer delay and avoid buffer overflow. This frame-skipping, however, produces undesirable motion discontinuity in the encoded video sequence. Conversely, if the buffer level is too low, there may be periods of time in which no bits are transmitted through the channel, and hence some channel bandwidth is wasted.

To overcome these two problems, a joint buffer control is usually used to maintain a buffer occupancy of about 50% of the buffer size after coding each frame. In order to do this, heuristic methods are usually employed, in which the target bit rate is increased if the current buffer level is less than half of the buffer size, and the target bit rate is decreased if the current buffer level is more than half of the current buffer size. Such schemes are disclosed in "Scalable Rate Control for MPEG-4 Video", H.J. Lee, T.H. Chiang and Y.Q. Zhang, IEEE Trans. Circuit Syst. Video Technol., 10:878-894, 2000, and in "MPEG-4 rate control for multiple video objects", A. Vetro, H. Sun and Y. Wang, IEEE Trans. Circuit Syst. Video Technol., 9:186-199, 1999. These schemes either encode video at a predefined fixed rate or at a predefined small set of fixed rates.

The existing schemes have problems when used in for example Internet applications and the streaming of video over the Internet. Due to the connectionless nature of the current Internet protocols and the routing mechanisms involved, the instantaneous bandwidth available to a particular user can vary widely in time and cannot in practice be previously known. The existing bit rate control schemes cannot adapt themselves quickly enough to the variations of channel bandwidth, and are not effective enough to achieve the objectives of Video over the Internet.

An aim of the present invention is to provide a bit rate control which provides for better video quality, especially but not exclusively in Internet applications.

Summary of the Invention

Viewed from a first aspect, the present invention provides a bit rate control system for the encoding of video data in which the encoded bits are placed in a buffer prior to transmission, and in which a target encoding bit rate is determined based on the fullness of the buffer, characterized in that the buffer is modelled by a fluid-flow traffic model preferably of the form:

$$B_c(n+1) = \max\{0, B_c(n) + T(n) - u(n)\}$$

where $B_c(n)$ denotes the buffer level at time n;
$T(n)$ is the actual encoding bit rate; and
$u(n)$ is the channel output rate.

The system of the present invention is able to keep the buffer occupancy closer to its target, which is preferably set at a predefined percentage (preferably about 50%) of a safety margin used to determine whether a frame of the video sequence to be encoded should be skipped, and to adapt itself faster to the variations of the channel bandwidth, and so will skip fewer frames at a low

bandwidth. This therefore provides a higher overall video quality, and is attractive for video over the Internet.

Preferably, the target encoding bit rate is given by the equation:

$$\tilde{f}(n) = A + (1-\gamma)\frac{\tilde{\gamma} * B_s}{2} + (\gamma - 1)B_c(n)$$

where A is the channel output rate;

$\tilde{\gamma}$ is a buffer safety margin;

B_s is the buffer size;

$B_c(n)$ is the current buffer level; and

$0 \leq \gamma < 1$ is an adjustable parameter.

"A" may be equal to the number of bits available for encoding all of the inter-frames of a current group of frames being encoded divided by the number of inter-frames to be encoded in the current group of frames. Alternatively, when for example providing video over the Internet, "A" may be the actual bandwidth estimated by using the packet loss information. This allows the variation of the channel bandwidth to be directly incorporated into the buffer control, and allows the system to adapt itself in time.

Meanwhile, the target bit rate is preferably modified based on the remaining bits available for encoding and on the remaining frames to be encoded. It may thus be:

$$f(n) = \max\{\beta * \frac{T_r}{N_r} + (1-\beta) * \tilde{f}(n), \frac{T_r}{3N_r} + H_{hdr}(n-1)\}$$

where $0 < \beta < 1$ is an adjustable parameter;

T_r is the number of remaining bits available for encoding;

N_r is the number of frames remaining to be encoded; and

$H_{hdr}(n-1)$ is the amount of overhead bits used for the previous frame.

After the target bit rate is determined, a rate-distortion model preferably of the following form is further applied to determine the corresponding quantization parameter:

$$R = c_2 \frac{\sigma^2}{Q^2} + c_1 \frac{\sigma}{Q} + H_{hdr}$$

where R is the total number of bits used to encode a frame;

Q is the quantization parameter;

c_1 and c_2 are first and second order coefficients;

σ is an index of video coding complexity; and

H_{hdr} is the amount of overhead bits used.

Further preferably, the coefficients of the Rate-Distortion model are updated based upon data from a plurality of previous frames. The number of previous frame used is preferably determined by a sliding window mechanism, wherein the value of the current window size W(n) is given by:

$$W(n) = \min\{W(n-1)+1, \varsigma(n) * W_{\max}\}$$

where W_{max} is a preset constant;

$$\varsigma(n) = \min\left\{\frac{\sigma(n-1)}{\sigma(n)}, \frac{\sigma(n)}{\sigma(n-1)}\right\}; \text{ and}$$

$\sigma(n)$ is the maximum absolute difference of the frame at time n.

Such a sliding window mechanism smoothes the impact of scene changes, and changes the window size gradually.

After the current frame is encoded, the total number of actual bits used to encode the current frame is added to the current buffer level. If the buffer is in danger of overflow, a switched frame skipping mechanism is preferably used to compute the number of skipped frames.

In one frame skip control, after the current frame is encoded, the next frame to be encoded will be skipped, if:

$$B_c(n+1)+T(n)-A \geq B_s * \tilde{\gamma} + T_B(n)$$

where $B_c(n+1)$ is the current buffer level;

$T(n)$ is the actual number of bits used to encode the current frame;

A is the channel output rate;

B_s is the buffer size;

$\tilde{\gamma}$ is a pre-determined buffer safety margin;

$$T_B^2(n)=\frac{1}{W(n)}\sum_{j=1}^{W(n)}\left(T(S(j,n))-\frac{1}{W(n)}\sum_{j=1}^{W(n)}T(S(j,n)) \right)^2$$

and $T(S(j,n))$ $(1 \leq j \leq W(n))$ denotes the total number of actual bits generated in the encoding of the previous W(n) frames.

In an alternative frame skip control, a frame skipping parameter N_{post} is set to skip the next N_{post} frames so that the following buffer condition is satisfied:

$$B_c(n+1) < \tilde{\gamma} B_s$$

where

$$B_c(n+1) = \max\{0, B_c(n)+T(n)-A(N_{post}+1)\}$$

$B_c(n)$ is the buffer level at time n;

$T(n)$ is the actual number of bits used to encode the current frame;

A is the channel output rate;

B_s is the buffer size; and

$\tilde{\gamma}$ is a pre-determined buffer safety margin.

The first-mentioned skipping control is preferably provided as a predictive switching control, the second-mentioned skipping control is preferably provided as a post-frame skipping control, and the skipping controls are preferably switched between one another based on the following switching law:

a) The predictive frame skipping control is switched to the post-skipping control if a frame is skipped; and

b) The post-skipping control is switched to the predictive frame skipping control if the current frame is not skipped.

The present invention also extends to a method for the encoding of a video sequence in accordance with the above system features, and to computer software for implementing the above system and method features.

It further extends to the use of the above features independently of one another, with for example the Rate-Distortion model defined above being in itself a new and advantageous model for use in bit rate control.

Brief Description of the Drawings

The present invention will hereinafter be described in greater detail by reference to the attached drawings which show an example form of the invention. It is to be understood that the particularity of the drawings does not supersede the generality of the preceding description of the invention.

Figure 1 is a diagram of the structure of a typical network over which video streaming may be provided; and

Figure 2 is a functional block diagram of a video encoder scheme according to an embodiment of the present invention.

Detailed Description of the Invention

Fig. 1 shows a typical Internet structure over which a video sequence may need to be transmitted from a source 1 to one or more receivers 2. Due to the amount of data in a video sequence, the data must be compressed, otherwise the required transmission bit-rate would be unachievably high.

Thus, an encoder 3 is provided at the source 1 in order to compress the video data, and decoders 4 are provided at the receivers 2 in order to decode the data and reconstruct the video sequence. In between the encoder 1 and decoders 4, the compressed data is routed through various servers 5 and over what may be many different types of transmission channel 6.

Various different encoding systems have been provided for the compression of video data, and, for example, MPEG video compression is often employed. The current MPEG standards are MPEG-1 and MPEG-2, which are similar in basic concept, and MPEG-4 which is able to provide a low-bandwidth multimedia format that can contain a mix of media (including recorded video images and sounds and their computer-generated counterparts), and uses the concept of "Video Objects" to transmit independent images of arbitrary shape.

In MPEG compression, a video sequence is broken into a number of Groups of Pictures (GOP), each of which comprises a number of picture frames. Each frame is broken into a series of slices, and each slice consists of a set of macroblocks comprising arrays of luminance pixels and associated chrominance pixels. The macroblocks are divided into 8x8 blocks for encoding. Each block undergoes a Discrete Cosine Transform (DCT) to provide an array of DCT coefficients that are then quantized to force various of the coefficients (generally higher frequency coefficients) to zero so as to reduce the amount of data to be transmitted. Quantization is carried out by multiplying the DCT coefficient array by a quantization matrix, each value in the matrix being scaled by a quantization parameter. The matrix and quantization parameter can be altered on a frame-by-frame and/or block-by-block basis to alter the amount of

compression. The quantized coefficients then undergo further encoding to compress the transmission data still further.

The frames in a GOP comprise an Intra-frame (I frame) that is spatially compressed (in accordance with the above method), and Inter-frames (P and/or B frames) that are also temporally compressed in a motion-compensated prediction manner. Thus, each P frame in a sequence is predicted from the frame immediately preceding it, and each B frame is predicted from preceding and succeeding frames.

MPEG-4 also includes a Video Object layer between the frame layer and macroblock layer for specifying different independent objects within a scene.

In order to optimise video quality over a bit-rate range, e.g. in video-streaming to a number of receivers having different bandwidth capabilities, MPEG-4 also provides a Fine Granularity Scalability (FGS) scheme in which the coding of the video data is provided by a base layer and an enhancement layer, the base layer being designed to meet the lower bound of the bit rate range and the enhancement layer meeting the upper bound of the bit-rate range. The base layer is coded as discussed above, and the enhancement layer takes the original and reconstructed DCT coefficients of the base layer, and subtracts the reconstructed coefficients from the originals to provide a residue that is then encoded and transmitted with the base layer. The receivers of the data decode the base layer to provide a video signal based on the lowest bit rate range, and can improve the quality by decoding various amounts of the enhancement layer.

The present invention relates to a bit rate control scheme for the compression of video data, and may for example be used in encoding the base layer of an FGS scheme. It may especially be used in the FGS disclosed in the co-pending International PCT patent application filed in Singapore on 25 May 2001 and entitled “A Fine Granularity Scalability Scheme”.

The present bit-rate control scheme consists of three layers, namely the GOP layer, the frame layer and the video object layer. The whole scheme is shown in Fig. 2.

The GOP layer rate control 1 is used to allocate bits to each GOP of the video sequence, each GOP being composed of one I frame and a number of P and B frames.

The total number of bits available for the video sequence will be:

TB = x R

where is the duration of the video sequence; and
R is the bit rate for the sequence.

Assuming that the total number of I frames is $\hat{N}_I$ and that the number of P and B frames in the ith GOP are $\hat{N}_{P,i}$ and $\hat{N}_{B,i}$, and that the frames have weightings of W_I, W_P and W_B, then the number of bits allocated to the ith GOP is:

$$TB_i = TB * \frac{\hat{N}_{P,i}W_P + \hat{N}_{B,i}W_B + W_I}{\sum_{i=1}^{\hat{N}_I}(\hat{N}_{P,i}W_P + \hat{N}_{B,i}W_B + W_I)}$$

For the sake of the present embodiment and for simplicity, it is assumed that each GOP has the same structure, and so the GOP Layer Rate Control will allocate each GOP the following number of bits:

$$TB_i = \frac{TB}{\hat{N}_I}.$$

After the GOP layer rate control at block 1, the encoder carries out a buffer initialization at block 2, conducts the Intra-coding of the I-frame at block 3, updates a Rate-Distortion model at block 4 and checks as to whether the next

frame must be skipped at a skip-frame block 5 (e.g. because of possible buffer overrun).

Inter-coding is then performed in which the encoder 3 performs a joint buffer control at block 6, a Frame Layer Target Bit Rate calculation at block 7 and a Quantization Parameter calculation at block 8, before carrying out the Inter-coding of the P or B frame at block 9. After encoding of the frame, the R-D model update and Frame-skip control are again carried out at blocks 4 and 5 before conducting the encoding of the next inter-frame through block 6, etc.

Where the encoder scheme is used in the Video Object layer, the encoder also conducts a Target Bit Rate Allocation at block 10, and calculates a shape threshold in block 8 along with the quantization parameter calculation.

The part of the bit rate control in the frame layer consists of three stages: the initialization, pre-encoding and post-encoding stages.

(a) Initialization Stage

In the initialization stage of block 2, the encoder carries out three main tasks with respect to the frame layer control, these being:

(i) initialization of the buffer size based on latency requirements;

(ii) subtraction of the bit count of the I-frame from the bit count of the ith GOP; and

(iii) initialization of the buffer fullness – If the first GOP is encoded, then buffer fullness is set at 50% of a buffer safety margin (which will be 40% of the buffer size assuming a safety margin of 80%). Otherwise, the buffer fullness is set at the end level of the previous GPO.

The I-frame is quantized using an initial quantization value of Q_o. The remaining available bits $R_0(i)$ for encoding all of the subsequent inter-frames can be calculated as:

$$R_0(i) = TB_i - \kappa_i + (0.5 * B_s * \tilde{\gamma} - \hat{B}_c(i))$$

where TB_i is the number of bits available to encode the ith group of frames;

κ_i is the number of bits used to encode the ith intra-frame;

B_s is the buffer size;

$\tilde{\gamma}$ is the buffer safety margin for skipping frames, having a typical value of 0.8; and

$\hat{B}_c(i)$ is the buffer level at the start of encoding of the ith group of frames, with $\hat{B}_c(1) = 0.5 * B_s * \tilde{\gamma}$.

The channel output rate (the average number of bits to be drained from the buffer per frame encoding) is then $R_0(i) / \hat{N}_{P,i}$.

(b) Pre-encoding Stage

The pre-encoding stage includes setting a target bit rate for the encoding of the next video frame in the GOP, and setting the quantization parameter for quantization of the DCT coefficients in accordance with the target bit rate.

When the number of bits in the buffer is too large (e.g. is predicted to exceed a safety margin), the encoder usually skips some frames to reduce the buffer delay and avoid buffer overflow. This however produces undesirable motion discontinuities in the encoded video sequence. Conversely, if the buffer level is too low, there may be periods of time in which no bits are transmitted through the channel, and channel bandwidth is wasted.

In order to overcome these problems, a frame level control is adopted which sets the target bit rate so as to attempt to maintain a buffer occupancy after the coding of each frame of about 50% of the buffer safety margin (i.e. about 40% of the buffer size for a 0.8 safety margin).

It should be noted that this differs from the prior art, which sets the target buffer fullness at the middle level of the buffer. The present scheme enables a low encoder buffer delay to be maintained and the total delay to be reduced.

In order to determine the target bit rate, the dynamics of the buffer are represented by a fluid-flow traffic model with $B_c(n)$ denoting the buffer level at time n:

$$B_c(n+1) = \max\{0, B_c(n) + T(n) - u(n)\} \qquad (1)$$

where T(n) is the actual encoding bit rate; and
u(n) is the channel output rate.

Using equation (1) and linear system control theory (see for example Chi-Tsong Chen, "Linear system theory and design", Rinehard and Winston, New York, 1984), the target bit rate is scaled based on the buffer size B_s, the current buffer level $B_c(n)$ and the channel output rate $R_0(i)/\hat{N}_{P,i}$, and is given by:

$$\tilde{f}(n) = \max\left\{0, \frac{R_0}{\hat{N}_{P,i}} + (1-\gamma)\frac{\tilde{\gamma} * B_s}{2} + (\gamma - 1)B_c(n)\right\}$$

where $0 \le \gamma < 1$ is an adjustable parameter having a typical value of 0.75.

When calculating the bit rate for the frame, the number of remaining bits T_r allocated to the current GOP and the remaining number of frames N_r of the current GOP should also be taken into account to ensure that there are available bits for the remaining frames, and so the final frame bit rate is:

$$f(n) = \max\{\beta * \frac{T_r}{N_r} + (1-\beta) * \tilde{f}(n), \frac{T_r}{3N_r} + H_{hdr}(n-1)\}$$

where $0 < \beta < 1$ is an adjustable parameter having a typical value of 0.585; and
$H_{hdr}(n-1)$ is the amount of bits used for overhead data, that is, the bits used for non-texture data, e.g. shape information, motion vector information and header information.

It should be noted that the above method of using a fluid-flow model departs from the prior art use of heuristic methods for determining the target bit rate, and enables the buffer occupancy to be kept much closer to the target, so that fewer frames are skipped.

The present model-based method may be used in any suitable video transmission system, and is especially attractive when MPEG-4 video is transported over the Internet where variations in bandwidth occur. Using the heuristic approach, adjustment of the joint buffer control has a delay of one step, and cannot adapt itself in time to the variations in channel bandwidth. However, with the present model-based method, when the channel bandwidth is time-varying, the term $R_0(i)/\hat{N}_{P,i}$ may be replaced by the estimated actual channel bandwidth, e.g. by using the packet loss information. Thus, the variation of the channel bandwidth can be incorporated into the present joint buffer control, and the scheme can adapt itself in time.

A further point to note is that the receiver synchronization of a continuous media stream must deal with delay differences and variations. Since the present frame-layer control keeps the buffer occupancy much closer to the target (50% of the safety margin (40% of the buffer size)), the playout buffer delay can be reduced, and so the total delay is further reduced.

Once the target bit rate is determined, the corresponding quantization parameter, Q, can be computed by using a Rate-Distortion model, which takes the form of the following quadratic model:

$$R = c_2 \frac{\sigma^2}{Q^2} + c_1 \frac{\sigma}{Q} + H_{hdr}$$

where R is the total number of bits used to encode a frame;

Q is the quantization parameter;

c_1 and c_2 are first and second order coefficients;

σ is the mean absolute difference of texture computed using the motion-compensated residual for the luminance component (an index of video coding complexity); and

H_{hdr} is the amount of bits used for overhead data, that is, non-texture data, e.g. video/frame syntax, bits used for shape information, motion vector information and header information.

(c) Post-encoding Stage

The post-encoding stage includes the processes of updating the parameters c_1 and c_2 of the Rate-Distortion model and determining whether any frame-skipping is necessary to prevent possible buffer overflow.

The statistics of quantization parameter value and bit rate value, taken from a number of previously encoded frames including the immediately preceding frame, are used to provide improved parameters c_1 and c_2 for the R-D model by using a linear regression technique.

The number of frames to use is based on a sliding window mechanism, which is designed to smooth the impact that a scene change might have in the updating of the R-D model.

If the complexity changes significantly, i.e. in high motion scenes, a smaller window with more recent data points after the change is used. Otherwise, a window with more data points is used. To ensure that the window size is not varied too rapidly, the window size is increased gradually.

Thus, the value of the current window size W(n) is given by:

$$W(n) = \min\{W(n-1)+1, \varsigma(n) * Max\_Sliding\_Window\}$$

where Max_Sliding_Window is a preset constant, and may be set to e.g. 20; and

$$\varsigma(n) = \min\left\{\frac{\sigma(n-1)}{\sigma(n)}, \frac{\sigma(n)}{\sigma(n-1)}\right\}.$$

The selected sample data points within the window W(n) are denoted as S(j,n) $(1 \leq j \leq W(n))$.

For the selected data points, the encoder collects the quantization parameter statistics Q(j) and the actual bit rate statistics T(j), and, using a linear regression technique, the parameters can be obtained by:

$$c_2 = \frac{c_3 - c_4}{W(n)\sum_{j=1}^{W(n)} \frac{\sigma^2(S(j,n))}{Q^2(S(j,n))} - \left(\sum_{j=1}^{W(n)} \frac{Q(S(j,n))}{\sigma(S(j,n))}\right)^2}$$

$$c_1 = \frac{\sum_{j=1}^{W(n)}\left[\left(T(S(j,n)) - H_{hdr}(S(j,n))\right)\frac{Q(S(j,n))}{\sigma(S(j,n))} - \frac{\sigma(S(j,n))}{Q(S(j,n))}c_2\right]}{W(n)}$$

$$c_3 = W(n)\sum_{j=1}^{W(n)}\left(T(S(j,n)) - H_{hdr}(S(j,n))\right)$$

$$c_4 = \sum_{j=1}^{W(n)}\left(T(S(j,n)) - H_{hdr}(S(j,n))\right)\frac{Q(S(j,n))}{\sigma(S(j,n))}\sum_{j=1}^{W(n)}\frac{\sigma(S(j,n))}{Q(S(i,n))}$$

After updating the R-D model, the total number of actual bits T(n) used to encode the current frame is added to the current buffer level, and a switched frame skip control is performed to prevent buffer overflow and overcome continuous frame skipping. The switched frame skipping control is composed of two basic controllers (a predictive frame skipping controller and a post frame skipping controller) and a corresponding switching law to determine the active controller.

In the predictive frame skip controller, a function T_B is defined by:

$$T_B^2(n) = \frac{1}{W(n)} \sum_{j=1}^{W(n)} \left(T(S(j,n)) - 1 \frac{1}{W(n)} \sum_{j=1}^{W(n)} T(S(j,n)) \right)^2$$

where T(S(j,n)) ($1 \leq j \leq W(n)$ denotes the total number of actual bits generated in the encoding of the previous W(n) frames.

The next frame to be encoded will be skipped, if the current buffer level plus the estimated number of bits for the next frame is larger than the sum of $T_B(n)$ and some pre-determined threshold, called the safety margin, that is if:

$$B_c(n+1) + T(n) - A \geq B_s * \tilde{\gamma} + T_B(n)$$

where $B_c(n+1)$ is the current buffer level;

T(n) is the actual number of bits used to encode the current frame;

A is the channel output rate (which may be $R_0(i) / \hat{N}_{P,i}$ or is replaced by the estimated actual channel bandwidth);

B_s is the buffer size; and

$\tilde{\gamma}$ is the pre-determined safety margin.

If skipping takes place, the current buffer level is reduced by the channel output rate.

In the post frame skipping controller, a frame skipping parameter N_{post} is increased from zero until the following buffer condition is satisfied, the next N_{post} frames are then skipped by the encoder:

$$B_c(n+1) < \tilde{\gamma} B_s$$

where

$$B_c(n+1) = \max\{0, B_c(n) + T(n) - A(N_{post} + 1)\} .$$

The predictive frame skipping control is initially used, and the switching law is:

a) The predictive frame skipping control is switched to the post-skipping control if a frame is skipped; and

b) The post-skipping control is switched to the predictive frame skipping control if the current frame is not skipped.

Instead of using the switched frame-skipping control, the predictive or post frame skipping control may be used by itself.

Besides using the present method on the frame layer rate control, the above method may also be used to control the video object layer rate control.

In the video object rate control, the total target bit rate (as found in the frame layer control) is allocated to each video object according to its coding complexity, size and perceptual importance. Thus, for a given target bit rate, the target bit rate for an object i is given by:

$$f_i(n) = \left(f(n) - H_{hdr}(n-1)\right)\frac{\rho_i * \sigma_i^2}{\sum_{k=1}^{N} \rho_k * \sigma_k^2}$$

$$+ H_{hdr}(n-1)\frac{\tau \sum_j \left(MOT_{ijx}(n) + MOT_{ijy}(n)\right) + (1-\tau)\rho_i}{\tau \sum_l \sum_j \left(MOT_{ljx}(n) + MOT_{ljy}(n)\right) + (1-\tau)\sum_j \rho_j}$$

where ρ_i is the size of the video object i;

$$H_{hdr}(n-1) = \sum_{1=1}^{N} H_{hdr,i}(n-1) \ ;$$

$MOT_{ijx}(n)$ and $MOT_{ijy}(n)$ are the absolute values of the jth motion vector component within the object i at the time n; and

τ is an adjustable parameter $0 < \tau < 1$.

Also, to avoid using excessive bits for motion and shape information instead of for texture, and to balance the bit usage without imposing additional noticeable

distortion, the shape threshold values can be set dynamically based on the previous coding information.

In the adaptive threshold shape control, let

$$H_B^2(i) = \frac{1}{W_i(n-1)} \sum_{j=1}^{W_i(n-1)} \left(H_{hdr,i}(S(j,n)) - \frac{1}{W_i(n-1)} \sum_{j=1}^{W_i(n-1)} H_{hdr,i}(S(j,n)) \right)^2$$

The threshold for the video object i, θ_i, is initially set to zero. if $f_i(n)$ is less than $H_{hdr,i}(n-1)$ - 1.25 $H_B(i)$ in the previous frame, then:

$$\theta_i = \min\{\theta_{\max}(i), \theta_i + \theta_{step}(i)\}.$$

where $\theta_{step}(i) > 0$ and $\theta_{max}(i) > 0$ are predefined.

If $f_i(n)$ is greater than $H_{hdr,i}(n-1) + 1.25 H_B(i)$, then it is decreased by:

$$\theta_i = \max\{\theta_{\max}(i), \theta_i - \theta_{step}(i)\}.$$

Otherwise, the threshold is not changed.

When controlling the video object layer, the switched frame skipping control will preferably be used.

Besides controlling the frame layer bit rate and the video object layer, the present scheme can also control the macroblock layer control. The method is thus scalable.

It is to be understood that various alterations additions and/or modifications may be made to the parts previously described without departing from the ambit of the invention, and that, in the light of the teachings of the present invention, the control scheme may be implement in software and/or hardware in a variety of manners.

Claims

1. A bit rate control system for the encoding of a video sequence in which encoded data is placed in a buffer prior to transmission, and in which a target encoding bit rate is determined based on the fullness of the buffer, characterised in that the buffer is modelled on a fluid-flow traffic model.

2. The system of claim 1, wherein said the fluid-flow traffic model is of the form:

$$B_c(n+1) = \max\{0, B_c(n) + T(n) - u(n)\} \quad ,$$

where $B_c(n)$ denotes the buffer level at time n;
$T(n)$ is the actual encoding bit rate; and
$u(n)$ is the channel output rate.

3. The system of claim 1, in which a rate-distortion model, used to compute a quantization parameter for the control system, has the form:

$$R = c_2 \frac{\sigma^2}{Q^2} + c_1 \frac{\sigma}{Q} + H_{hdr}$$

where R is the total number of bits used to encode a frame;
Q is the quantization parameter;
c_1 and c_2 are first and second order coefficients;
σ is an index of video coding complexity; and
H_{hdr} is the amount of overhead bits used.

4. The bit rate control system of claim 1, 2 or 3, wherein a buffer occupancy target is set at a predefined percentage of a safety margin, said safety margin being used to determine whether a frame of the video sequence to be encoded should be skipped.

5. The bit rate control system of claim 4, wherein said buffer target occupancy is set to about 50% of said safety margin.

6. The bit rate control system of any preceding claim, wherein said target encoding bit rate is given by the equation:

$$\tilde{f}(n) = A + (1-\gamma)\frac{\tilde{\gamma} * B_s}{2} + (\gamma - 1)B_c$$

where A is the channel output rate;

$\tilde{\gamma}$ is a buffer safety margin;

B_s is the buffer size;

$B_c(n)$ is the current buffer level; and

$0 \leq \gamma < 1$ is an adjustable parameter.

7. The bit rate control system of claim 6, wherein A is equal to the number of bits available for encoding all of the inter-frames of a current group of frames being encoded divided by the number of inter-frames to be encoded in the current group of frames.

8. The system of claim 7, wherein the available bits $R_0(i)$ for encoding the inter-frames of the ith group of frames is:

$$R_0(i) = TB_i - \kappa_i + (0.5 * B_s * \tilde{\gamma} - \hat{B}_c(i))$$

where TB_i is the number of bits available to encode the ith group of frames;

κ_i is the number of bits used to encode the ith intra-frame;

B_s is the buffer size;

$\tilde{\gamma}$ is the buffer safety margin; and

$\hat{B}_c(i)$ is the buffer level at the start of encoding the ith group of frames.

9. The bit rate control system of claim 6, wherein A is the estimated actual channel bandwidth.

10. The system of any preceding claim, wherein the target bit rate is modified based on the remaining bits available for encoding and on the remaining frames to be encoded.

11. The system of claim 10, wherein the target bit rate is:

$$f(n) = \max\{\beta * \frac{T_r}{N_r} + (1-\beta) * \tilde{f}(n), \frac{T_r}{3N_r} + H_{hdr}(n-1)\}$$

where $0 < \beta < 1$ is an adjustable parameter;

T_r is the number of remaining bits available for encoding;

N_r is the number of frames remaining to be encoded; and

$H_{hdr}(n-1)$ is the amount of overhead bits used for the previous frame.

12. The system of any preceding claim, wherein the bit rate control uses a rate-distortion model to determine the quantization parameter for a frame to be encoded, and wherein the coefficients of said model are updated based upon data from a plurality of previous frames, the number of previous frame used being determined by a sliding window mechanism, wherein the value of the current window size W(n) is given by:

$$W(n) = \min\{W(n-1)+1, \varsigma(n) * W_{\max}\}$$

where W_{max} is a preset constant; and

$$\varsigma(n) = \min\left\{\frac{\sigma(n-1)}{\sigma(n)}, \frac{\sigma(n)}{\sigma(n-1)}\right\}.$$

13. The system of any preceding claim, wherein after the current frame is encoded, the next frame to be encoded will be skipped, if:

$$B_c(n+1)+T(n)-A \geq B_s * \tilde{\gamma} + T_B(n)$$

where $B_c(n+1)$ is the current buffer level;

$T(n)$ is the actual number of bits used to encode the current frame;

A is the channel output rate;

B_s is the buffer size;

$\tilde{\gamma}$ is a pre-determined buffer safety margin; and

$$T_B^2(n)=\frac{1}{W(n)}\sum_{j=1}^{W(n)}(T(S(j,n))-\frac{1}{W(n)}\sum_{j=1}^{W(n)}T(S(j,n)))^2$$

where $T(S(j,n))$ $(1 \leq j \leq W(n))$ denotes the total number of actual bits generated in the encoding of the previous $W(n)$ frames.

14. The system of any of claims 1 to 13, wherein after the current frame is encoded, the total number of actual bits used to encode the current frame is added to the current buffer level, and wherein a frame skipping parameter N_{post} is set to skip the next N_{post} frames so that the following buffer condition is satisfied:

$$B_c(n+1) < \tilde{\gamma} B_s$$

where

$$B_c(n+1) = \max\{0, B_c(n)+T(n)-A(N_{post}+1)\}$$

where $B_c(n)$ is the buffer level at time n;

$T(n)$ is the actual number of bits used to encode the current frame;

A is the channel output rate;

B_s is the buffer size; and

$\tilde{\gamma}$ is a pre-determined buffer safety margin.

15. The system of claims 13 and 14, wherein the skipping control of claim 13 is provided as a predictive switching control, the skipping control of claim 14 is provided as a post-frame skipping control, and the skipping controls are switched between one another based on a switching law, said switching law being:

a) The predictive frame skipping control is switched to the post-skipping control if a frame is skipped; and

b) The post-frame skipping control is switched to the predictive frame skipping control if the current frame is not skipped.

16. A method for encoding a video sequence, including the step of placing encoded data into a buffer prior to transmission, and the step of determining a target encoding bit rate based on the fullness of the buffer, characterised by the step of modelling the buffer based on a fluid-flow traffic model.

17. The method of claim 16, wherein the fluid-flow model is of the form:

$$B_c(n+1) = \max\{0, B_c(n) + T(n) - u(n)\}$$

where $B_c(n)$ denotes the buffer level at time n;
T(n) is the actual encoding bit rate; and
u(n) is the channel output rate.

18. The method of claim 16 or 17, including the step of determining a quantization parameter for encoding the data based on a rate-distortion equation having the form:

$$R = c_2 \frac{\sigma^2}{Q^2} + c_1 \frac{\sigma}{Q} + H_{hdr}$$

where R is the total number of bits used to encode a frame;
Q is the quantization parameter;
c_1 and c_2 are first and second order coefficients;
σ is an index of video coding complexity; and

H_{hdr} is the amount of overhead bits used.

19. Computer software for the encoding of a video sequence, wherein encoded data is placed in a buffer prior to its transmission, and wherein the computer software includes a component which determines a target encoding bit rate based on the fullness of the buffer, characterised in that the software includes a component for modelling the buffer based on a fluid-flow traffic model.

20. The software of claim 19, including a component for determining a quantization parameter for encoding the data based on a rate-distortion equation having the form:

$$R = c_2 \frac{\sigma^2}{Q^2} + c_1 \frac{\sigma}{Q} + H_{hdr}$$

where R is the total number of bits used to encode a frame;

Q is the quantization parameter;

c_1 and c_2 are first and second order coefficients;

σ is an index of video coding complexity; and

H_{hdr} is the amount of overhead bits used.

21. A bit rate control system for the encoding of video data, wherein a rate-distortion model is used to determine a quantization parameter to use in the encoding, and characterised in that the rate-distortion model has the form:

$$R = c_2 \frac{\sigma^2}{Q^2} + c_1 \frac{\sigma}{Q} + H_{hdr}$$

where R is the total number of bits used to encode a frame;

Q is the quantization parameter;

c_1 and c_2 are first and second order coefficients;

σ is an index of video coding complexity; and

H_{hdr} is the amount of overhead bits used.

22. A bit rate control system for the encoding of a video sequence in which encoded data is placed in a buffer prior to its transmission, and in which a target encoding bit rate is determined based on the fullness of the buffer, characterised in that a buffer occupancy target is set at a set percentage of a safety margin, said safety margin being used to determine whether a frame of the video sequence to be encoded should be skipped.

23. The bit rate control system of claim 22, wherein said buffer target occupancy is set to about 50% of said safety margin.

24. A bit rate control system for the encoding of a video sequence in which encoded data is placed in a buffer prior to its transmission, and in which a target encoding bit rate is determined based on the fullness of the buffer, the bit rate control using a rate-distortion model to determine a quantization parameter for a frame to be encoded, and wherein the coefficients of said model are updated based upon data from a plurality of previous frames, the number of previous frame used being determined by a sliding window mechanism, characterised in that the value of the current window size W(n) is given by:

$$W(n) = \{\min W(n-1) + 1, \varsigma(n) * W_{\max}\}$$

where W_{max} is a preset constant; and

$$\varsigma(n) = \min\left\{\frac{\sigma(n-1)}{\sigma(n)}, \frac{\sigma(n)}{\sigma(n-1)}\right\}$$

$\sigma(n)$ being an index of video coding complexity.

25. A bit rate control system for the encoding of a video sequence in which encoded data is placed in a buffer prior to its transmission, and in which video data to be encoded is skipped if it is determined that buffer overflow may occur, characterised in that said skip control comprises:

a) a predictive skip control, in which, after the current frame is encoded, the next frame to be encoded will be skipped, if:

$$B_c(n+1)+T(n)-A \geq B_s * \tilde{\gamma} + T_B(n)$$

where $B_c(n+1)$ is the current buffer level;

$T(n)$ is the actual number of bits used to encode the current frame;

A is the channel output rate;

B_s is the buffer size;

$\tilde{\gamma}$ is a pre-determined buffer safety margin; and

$$T_B^2(n)=\frac{1}{W(n)}\sum_{j=1}^{W(n)}\left(T(S(j,n))-\frac{1}{W(n)}\sum_{j=1}^{W(n)}T(S(j,n))\right)^2$$

where $T(S(j,n))$ $(1 \leq j \leq W_s(n))$ denotes the total number of actual bits generated in the encoding of the previous $W(n)$ frames;

b) a post frame skip control in which after the current frame is encoded, the total number of actual bits used to encode the current frame is added to the current buffer level, and wherein a frame skipping parameter N_{post} is set to skip the next N_{post} frames so that the following buffer condition is satisfied:

$$B_c(n+1) < \tilde{\gamma} B_s$$

where

$$B_c(n+1) = \max\left\{0, B_c(n)+T(n)-A(N_{post}+1)\right\}$$

where $B_c(n)$ is the buffer level at time n;

$T(n)$ is the actual number of bits used to encode the current frame;

A is the channel output rate;

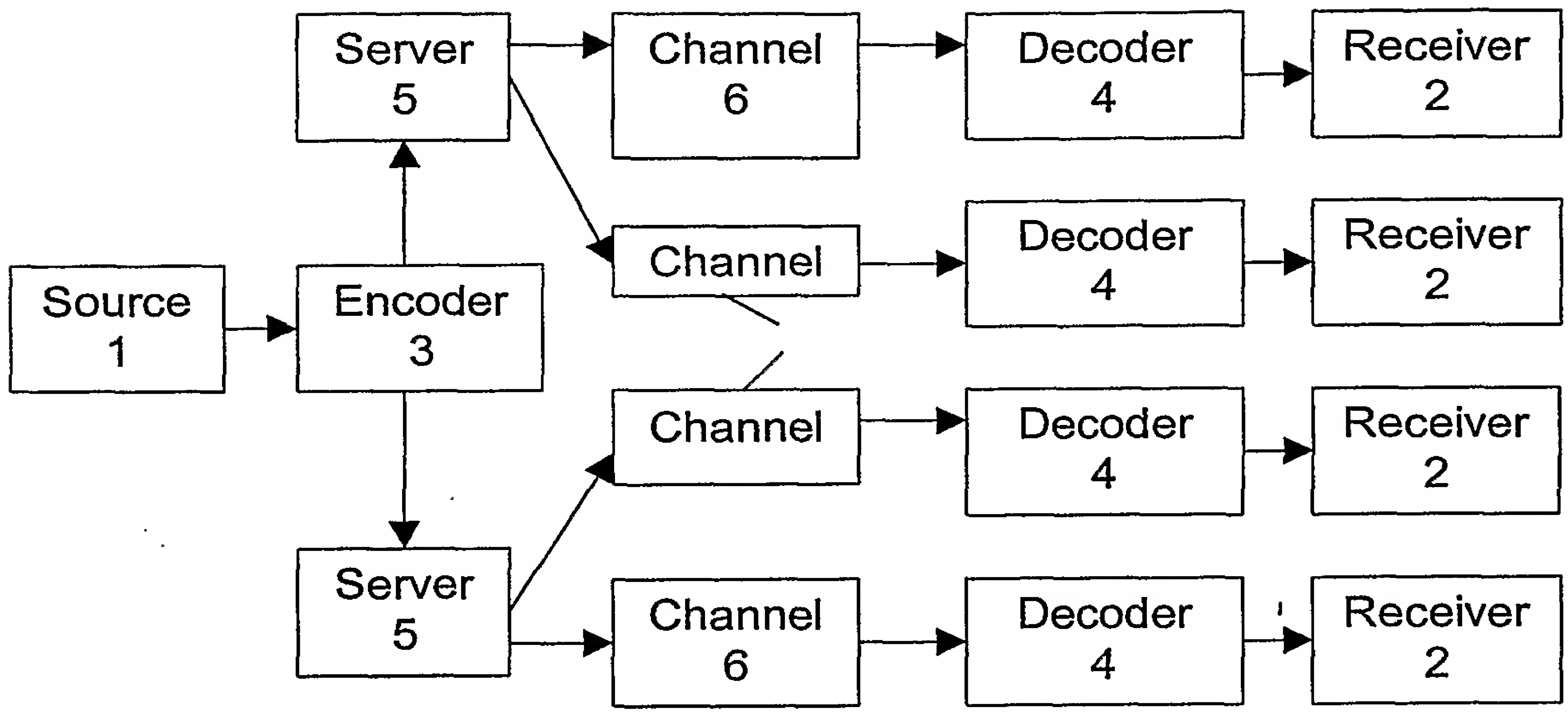
B_s is the buffer size; and

$\tilde{\gamma}$ is a pre-determined buffer safety margin.

c) a switching law through which the skipping controls are switched between one another, said switching law being:

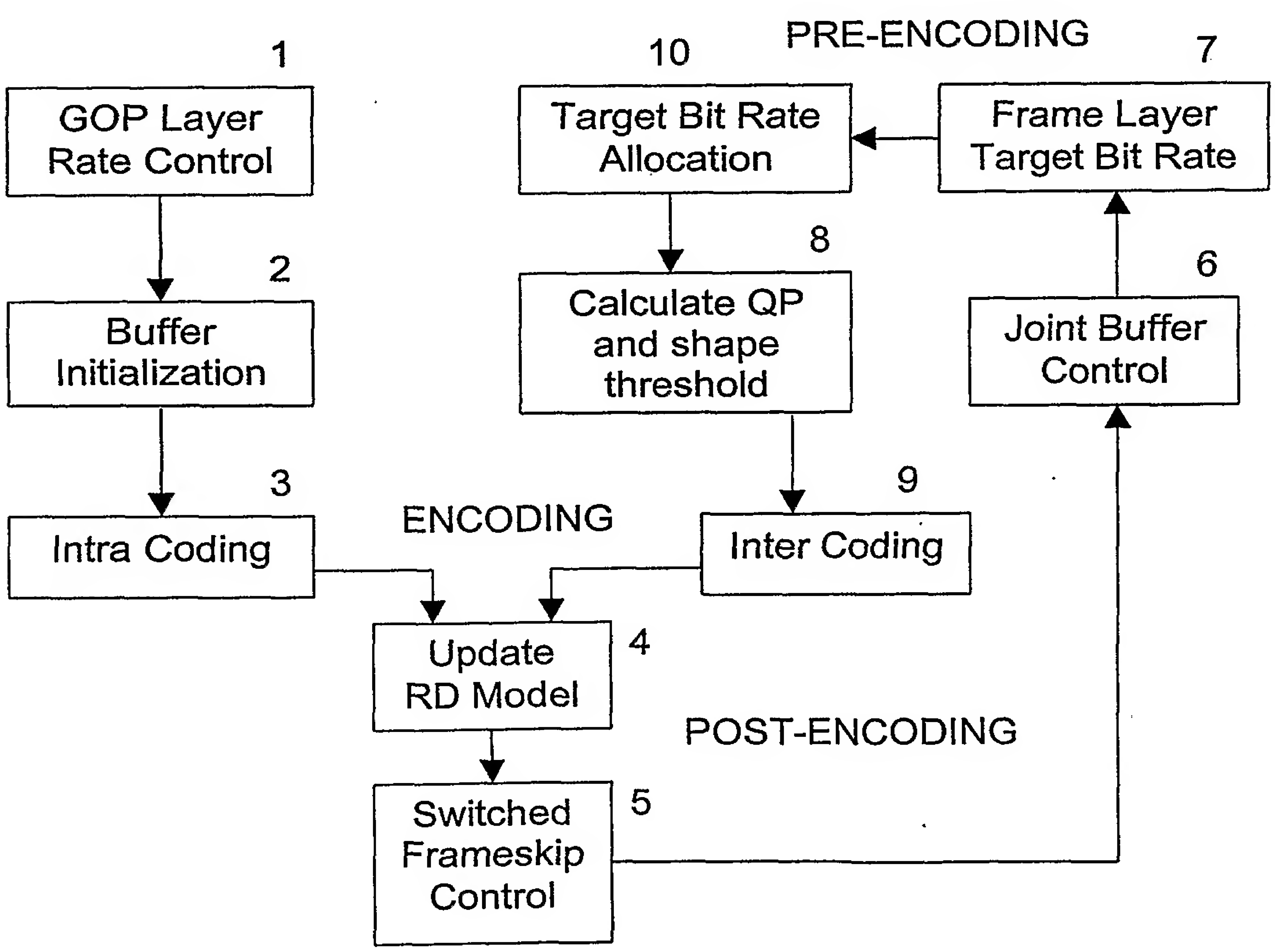
a) The predictive frame skipping control is switched to the post-skipping control if a frame is skipped; and

b) The post-frame skipping control is switched to the predictive frame skipping control if the current frame is not skipped.



Server 5
Channel 6
Decoder 4
Receiver 2
Channel
Decoder 4
Receiver 2
Source 1
Encoder 3
Channel
Decoder 4
Receiver 2
Server 5
Channel 6
Decoder 4
Receiver 2

Fig. 1



PRE-ENCODING
1
GOP Layer Rate Control
10
Target Bit Rate Allocation
7
Frame Layer Target Bit Rate
2
Buffer Initialization
8
Calculate QP and shape threshold
6
Joint Buffer Control
3
Intra Coding
ENCODING
9
Inter Coding
4
Update RD Model
POST-ENCODING
5
Switched Frameskip Control

Fig.2

INTERNATIONAL SEARCH REPORT

International Application No
PCT/SG 01/00105

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04N7/50 H04N7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 229 849 B1 (MIHARA KANJI) 8 May 2001 (2001-05-08)	1,2,16, 17,19
Y	column 11, line 7 - line 23	3-7,10, 12,13, 18,20
A	figure 6	8,9,11, 14,15, 21-25

	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 February 2002	01/03/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Georgiou, G

INTERNATIONAL SEARCH REPORT

Inter al Application No
PCT/SG 01/00105

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	LEE H-J ET AL: "SCALABLE RATE CONTROL FOR MPEG-4 VIDEO" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE INC. NEW YORK, US, vol. 10, no. 6, September 2000 (2000-09), pages 878-894, XP000959031 ISSN: 1051-8215 cited in the application	21-24
Y	paragraph ‘00II!; figure 2	3-7,10, 12,13, 18,20
A	paragraph ‘0III!	1,2,8,9, 11, 14-17, 19,25
X	VETRO A ET AL: "MPEG-4 RATE CONTROL FOR MULTIPLE VIDEO OBJECTS" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE INC. NEW YORK, US, vol. 9, no. 1, February 1999 (1999-02), pages 186-199, XP000802297 ISSN: 1051-8215 cited in the application paragraph ‘00II!	21-23
A	figure 1	1-20,24, 25
X	EP 0 892 555 A (MITSUBISHI ELECTRIC CORP) 20 January 1999 (1999-01-20)	21,22,25
A	page 4, line 22 -page 8, line 34; figure 2	1-20,23, 24
X	WO 98 37701 A (SARNOFF CORP) 27 August 1998 (1998-08-27)	21
A	page 9, line 1 - line 27; figure 2 page 14, line 25 -page 15, line 6	1-20, 22-25
X	LEE H-J ET AL: "RATE-DISTORTION BASED OPTIMIZATION FOR ZEROTREE ENTROPY WAVELET CODING" IEEE TRANSACTIONS ON CONSUMER ELECTRONICS, IEEE INC. NEW YORK, US, vol. 45, no. 3, August 1999 (1999-08), pages 650-660, XP000926980 ISSN: 0098-3063	21
A	paragraphs ‘0003!,‘0004!	1-20, 22-25

-/--

INTERNATIONAL SEARCH REPORT

Inte al Application No
PCT/SG 01/00105

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	GUANG-LIANG LI: "An analysis of transient loss performance impact of long-range dependence in ATM traffic" IEEE ATM WORKSHOP 1997. PROCEEDINGS LISBOA, PORTUGAL 25-28 MAY 1997, NEW YORK, NY, USA,IEEE, US, 25 May 1997 (1997-05-25), pages 603-610, XP010247447 ISBN: 0-7803-4196-1 paragraph ‘OIII! ---	1
A	GUSTAFSSON E ET AL: "Fluid traffic modelling in simulation of a call admission control scheme for ATM networks" MODELING, ANALYSIS, AND SIMULATION OF COMPUTER AND TELECOMMUNICATION SYSTEMS, 1997. MASCOTS ’97., PROCEEDINGS FIFTH INTERNATIONAL SYMPOSIUM ON HAIFA, ISRAEL 12-15 JAN. 1997, LOS ALAMITOS, CA, USA,IEEE COMPUT. SOC, US, 12 January 1997 (1997-01-12), pages 110-115, XP010211378 ISBN: 0-8186-7758-9 paragraph ‘0002! -----	1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter ıl Application No
PCT/SG 01/00105

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
US 6229849	B1	08-05-2001	GB	2337392	A	17-11-1999
			WO	9930504	A1	17-06-1999
			JP	11262008	A	24-09-1999
EP 0892555	A	20-01-1999	US	5969764	A	19-10-1999
			EP	0892555	A2	20-01-1999
WO 9837701	A	27-08-1998	CN	1247670	T	15-03-2000
			EP	0960532	A1	01-12-1999
			JP	2001512643	T	21-08-2001
			US	6243497	B1	05-06-2001
			WO	9837701	A1	27-08-1998